# Uitwerking Tentamen Kansrekening en Statistiek (153008) voor INF en TEL, donderdag 5 april 2007, 13.30-16.30 uur

**Opgave 1**

a. Niet waar: geldt alleen als de gebeurtenissen "man" en "van allochtone afkomst" onderling onafhankelijk zijn.

b. Niet waar: Als $P(A) > 0$ en $P(B) > 0$, dan zijn A en B afhankelijk want $0 = P(AB) \neq P(A)P(B)$.

c. Niet waar: de inkomens X en Y van een man en zijn vrouw zijn waarschijnlijk afhankelijk, dus geldt **niet** dat $var(X+Y) = var(X) + var(Y)$ (en ook de normaliteit van X+Y staat niet vast).

d. Niet waar: die bewering geldt voor een voorspellingsinterval.

e. Waar: bijvoorbeeld de intervalbreedte $2cs/\sqrt{n}$ (normaal model onbekende $\sigma$) wordt groter omdat c groter wordt bij toenemende betrouwbaarheid.

f. Waar: (zie e) de breedte wordt gehalveerd: $2cs/\sqrt{4n} = cs/\sqrt{n}$

g. Waar: grotere steekproefomvang leidt tot een groter onderscheidend vermogen (en een kleinere kans op een fout van de tweede soort), omdat de standaardafwijking van de steekproeffractie kleiner wordt bij toenemende n.

**Opgave 2**

a. De kansverdeling van X: $P(X=0) = 0.4$ en $P(X=1) = P(X=2) = 0.3$
$E(X) = \sum xP(X=x) = 0 + 1\times 0.3 + 2\times 0.3 = 0.9$
$E(X^2) = \sum x^2P(X=x) = 0 + 1\times 0.3 + 4\times 0.3 = 1.5$
$Var(X) = E(X^2) - E(X)^2 = 1.5 - 0.81 = 0.69$

b. Y heeft dezelfde verdeling als X, dus $E(X) = 0.9$ en $var(X) = 0.69$
$E(XY) = \sum\sum xyP(X=x \text{ en } Y=y) = 1\times 1\times 0.2 + 1\times 2\times 0.05 + 2\times 1\times 0.05+ 2\times 2\times 0.05 = 0.6$
$\rho(X, Y) = cov(X,Y)/ (\sigma_X\sigma_Y) = (0.6 - 0.9\times 0.9) / 0.69 \approx -0.304$
X en Y zijn dus licht negatief gecorreleerd.

c. $P(X+Y = 0) = 0.15$, $P(X+Y = 1) = 0.05+0.05 = 0.10$, $P(X+Y = 2)= 0.60$, $P(X+Y = 3) = 0.10$ en $P(X+Y = 4) = 0.05$.

d. $P(X=0|Y=2) = 0.20/0.30 = 2/3$ en $P(X=1|Y=2) = P(X=2|Y=2) = 0.05/0.30 = 1/6$
Dus $E(X|Y=2) = 0\times 2/3 + 1\times 1/6 + 2\times 1/6 = ½$

**Opgave 3**

a. Bereken a. $P(X \geq 3) = 1- P(X \leq 2) = 0.323$ (zie tabel)

b. Y = "het aantal telefoontjes in 4 minuten" is Poisson verdeeld met $\mu = 4\times 2 = 8$
$P(Y \geq 7) = 1- P(X \leq 6) = 0.313$ (zie tabel met $\mu = 8$).

c. $S = \sum X_i$ is bij benadering $N(\mu, \sigma^2)$ met $\mu = 60\times 2 =120$ en $\sigma^2 = 60 \times 2 = 120$
$P(S \geq 105) = P(S \geq 105) = P( Z \geq (105 - 120)/\sqrt{120} ) \approx P(Z \geq -1.37) = P(Z \leq 1.37) = 91.47\%$
(eventueel met continuïteitscorrectie, dus $P(S \geq 104.5) = \ldots= 92.14\%$)

**Opgave 4**

a. $E(X) = 1/\lambda_1 = 2$
$P(X > 2\mu_X \mid X > \mu_X ) = P( X > \mu_X )$ (wegens geheugenloosheid)
$= P(X > 2) = \int_2^\infty \frac{1}{2}e^{-x/2}dx = -e^{-x/2}\Big|_2^\infty = e^{-1} \approx 36.8\%$

b. $f_V(v) = \int_0^v \frac{1}{2}e^{-x/2}e^{-(v-x)}dx = e^{-v}e^{x/2}\Big|_0^v = e^{-v/2} - e^{-v}$ , voor $v \geq 0$ (en $f_V(v) = 0$ als $v < 0$)

c. $E(V) = E(X_1) + E(X_2) = 2 +1 = 3$
$var(V) = var(X_1) + var(X_2) = 4 +1 = 5$ wegens o.o.-heid van $X_1$ en $X_2$ (en $var(X) = 1/\lambda^2$ )

d. $Z = \max (X_1, X_2)$
$F(z) = P( \max (X_1, X_2) \leq z ) = P( X_1 \leq z \text{ en } X_2 \leq z) = P( X_1 \leq z)\times P(X_2 \leq z) = (1- e^{-z/2} )^2$ voor $z \geq 0$

**Opgave 5**

a. $E(X) = np$ en $var(X) = np(1-p)$

b. Voor $n = 400$ en $p = 0.01$ kan de Poisson verdeling met $\mu = 400\times 0.01 =4$ als benadering gebruikt worden. (voor een normale benadering is np te klein)
Voor $n = 400$ en $p = 0.10$ kan de N(40, 36)-verdeling als benaderende verdeling gebruikt worden.

c. Deze kans is $p^2$ en wordt geschat met $\hat{p}^2 = \frac{X^2}{n^2}$ ( $\hat{p}$ is de meest aannemelijke schatter van p)
Omdat $Var(X) = \mathbf{E(X^2)} - E(X)^2$ geldt:

$E(\hat{p}^2) = \mathbf{E(X^2)}/n^2 = [var(X) + E(X)^2]/n^2 = [np(1-p) + n^2 p^2]/n^2 = p^2 + p(1-p)/n$

Dus $E(\hat{p}^2) > p^2$, dus $\hat{p}^2$ is geen zuivere schatter van $p^2$.

d. Binomiaal model voor X, dus

$$90\% - BI(p) = \left(\hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right),$$

waarin $n$ =2500, $\hat{p} = \frac{x}{n} = \frac{1296}{2500} = 51.84\%$ en $c$ =1.645 uit de $N(0,1)$-tabel zodat $P(Z<c)= 0.95$.

Dus $90\% - BI(p) = (0.502, 0.535)$

e.

1. Model: X= het aantal personen die voor zijn in de steekproef van n =2500: X is B(2500, p) met onbekende p
2. Toets $H_0$: $p \le ½$ tegen $H_0$: met $\alpha \le 0.05$
3. Toetsingsgrootheid X is B(2500, ½ ) als p = ½ dus bij benadering N(1250, 625)
4. Waargenomen: X = 1296
5. Rechtseenzijdige toets: Als $X \ge c$, dan $H_0$ verwerpen
   $P(X \ge c \mid H_0) \le 0.05$ , dus $P([X -1250]/25 \ge [c- ½ -1250]/25 \mid H_0) \le 0.05$
   Uit de N(0,1)-tabel volgt dan dat $[c- ½ -1250]/25 \ge 1.645$, ofwel $c \ge 1291.6$, <u>dus c = 1292</u>
6. X= 1296 > 1292 = c , dus $H_0$ verwerpen
7. Met een onbetrouwbaarheid van 5% is aangetoond dat meer dan helft van de Nederlanders vóór het weren van de SP uit het bestuur is.